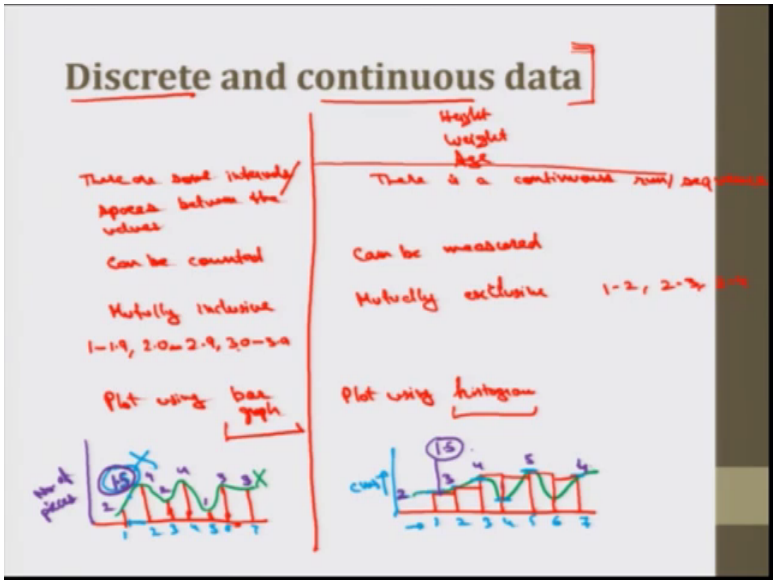
**Engineering Metrology**

**Prof. J. Ramkumar**
**Dr. Amandeep Singh Oberoi**
**Department of Mechanical Engineering**
**Department of Industrial & Production Engineering**
**Indian Institute of Technology, Kanpur**
**Indian National of Technology, Jalandhar**

**Lecture – 38**
**Discrete and continuous data**

So welcome back to the next part of the lecture.

(Refer Slide Time: 00:18)




So, next topic is Discrete data and continuous data, discrete is the term that is for the values which have nothing in between, continuous is when the curve is continuous, discrete events have it can just occur randomly. Continuous events when strains the heights the heights are continuous height weight age the time taken to complete an event, then the temperature which is this is these are all continuous kind of values. This is height, weight, age, temperature, then time; this is continuous data. But in discrete what we have? Discrete is the kind of categorical sometimes like is the tank full or empty, the continuous can be the capacity of the tank height is this tall or small. So, this is nothing between tall and small.

If I ok. If I can say very tall medium small, very small 5 categories but these are discrete the categorical variable is discrete, about where I can say over weight, then adequate

weight, then moderately weighted, adequately under weighted. Then underweight these 5 category they can divided into 2 3 4 5 whatever number I like this discrete events, if there is also the big study on discrete event simulation and talk about the events and variables in the probability distribution. Discrete event simulation when the events are discrete there is no specific time, but can come at any point of time I like randomly, so for that also simulation concepts are there.

So, for the discrete and continuous data discrete data takes only a particular value, there may be potentially be an infinite number of those values. But in the distinct (Refer Time: 00:18) I could say there is no grey area in between there is no connection between the 2 events or 2 values here in the discrete and continuous they are connected. So, discrete data can be numeric like discrete data examples if not categorical I can say the number of apples, the number of products those are produced by the factory in a specific number say in a whole day not; in whole day this number of products ok.

Then number of defects in welding in the specific welding, total number of defects in the final product total number of customers who are rejecting the product; these are the numbers ok. These are discrete numbers we cannot connect between customer number 5 and customer number 4 who has rejected that, this is the discrete information. Discrete means there are some intervals or species between the values, in continuous data the data falls on a cut in the sequence; there is a continuous run or sequence.

So, if I see the nature of the values for the discrete and continuous data, the continuous data values can be measured, but the discrete values cannot be measured. However, we can we can count the discrete value number of events the discrete events, so that we can count them but we cannot exactly measure them. So, these are if I say the nature this is this can be counted, this can be measured and these are mutually inclusive. These are mutually exclusive. So, the overlapping or mutual exclusive classification such as the intervals like between 1 to 2, between 2 to 3 test is done for continuous data these are mutually exclusive, I can say between this thing between 2 to 4. So, this can be done for the continuous data, but say this is opposed to this is something contrary to the non overlapping of the mutually inclusive classification.

Like the classification in discrete data could be because we cannot overlap them, we can just say I will put the numbers like 1 to 1.9 then 2.0 to 2.9, then 3.0 to 3.9. So these are

mutually inclusive discrete data to continuous data is mutually exclusive. So, the graph for the plot for the discrete data is generally the, per graph and for the continuous data it is generally the histogram. I can say the plot using per graphs, this is plot using histogram; you know this is data visualization. Data visualization is the science in statistics in which the graphical representation of data is done. The certain ways here you know what is what a bar graphs what histograms what a pie charts and what fan charts that these are actually a little complex one.

Then we are set another ways like line graph scatter diagram data visualisation is also important, which kind of data is they what is the nature of the data hock would be best represent the data into a graph. They are actually 3D graphs also available using set in software's in computers we can also draw 3 graph cells like limited to 3 dimensions only. Sometime they are multiple dimensions and that we have to draw the separate 2 dimension or 3 dimension graph for that.

So, this is also important you know when you do the measurements to just do some calculations and you have to represent the report finally, this is acceptable or this is not acceptable or what should be the final action that is to be taken, what has what is the final decision or what conclusion have withdrawn from that, for that it is important to draw the data properly. I will talk about data visualization in the forthcoming lectures as well we have to properly plot this data as well that is best suited to our results, it is said that one figure is equivalent to a 1000 words.

You write 1000 of 100 of for you write 1000 number of words, but you represent one figure showing various graphs ok. This is this one a calibre is behaving like this another company calibre is giving this reading, it just show the graph various bar charts or various lines in them they show which is selected you use colours for them that would talk everything.

The first thing when you read a research paper I think what I do after reading the abstract, the first thing I do is I look at the figures if it if they are possible figures and tables these illustrations are very important. So, just talking about the important plotting here for discrete it is the bar graph is used for the continuous data histogram is use, I will tell u how do we used number of histograms to draw a normal curve to draw a continuous curve ok, then there is no range between the highest and lowest point ok. If

the range is there what kind of line is drawn that is in the discrete data also sometimes the range is there in continuous data that is also there.

So, the major difference between the discrete and continue it is continuous data is that, in discrete data we have the spaces in between there is no connection between the 2 lines. There is 13 lines there is no connection between 2 lines we cannot draw where there 2 or 3 lines like this we cannot draw connect them, if the lines like this in discrete data and in continuous data also if I having some values like this it is not sensible to connect them like this is not recommended.

Yes, in continuous data if the values are there those can be connected, why it is not sensible can you think off, in the discrete data; why we cannot connect, why it is not recommended to connect here, because the space in between the space this space the blue line this does not mean anything. If I point something here if I say this is point 1 2 3 4 5 6 and 7 this is nothing known as 0.1 0.5 this does not exist the spaces have no meaning in discrete data.

In continuous data yes if there are certain points one 2 3 4 5 6 and 7, any 5 having if I am having a point here if this is a length ok. This is a length let me say 2 centimetre 3 centimetre 4 centimetre 1 cm 5 centimetre and again 4 centimetre these are the lengths here ok. Limits a 2 this is a value here I will put centimetres this value is 2 actually this 1 2 3 4 5 6 7 written I will change the colour, in the blue colour is actually the nominal scale this is just a number of the reading that number the note that count; but the name of the reading, reading 1 reading 2 reading 3 reading 4 and 5 and 7 in the blue colour.

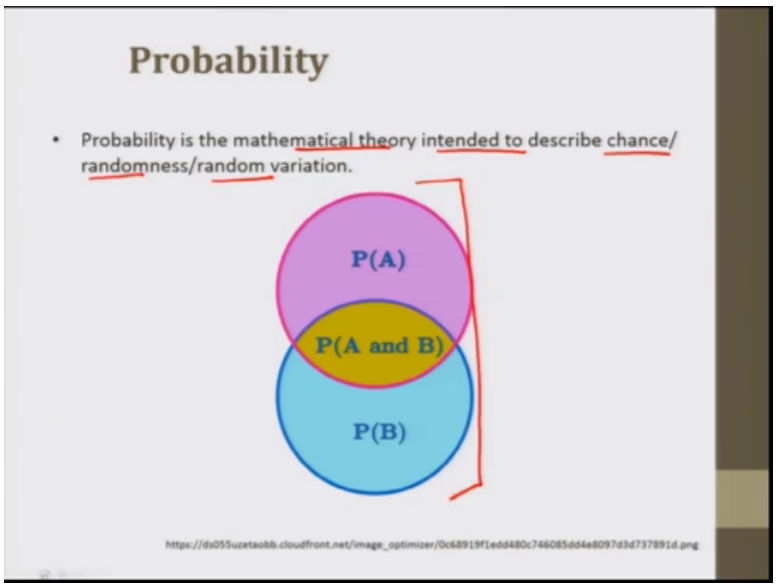
Now, I will pick the another colour this is I will pick purple colour this is actually the value 2 centimetres 3 centimetres 4 centimetres and 1 centimetre 5 centimetre and maybe this is again 4 centimetre ok. These are counts in discrete data one 2 3 4 5 6 7 actually the written in blue are only the original data, now in purple colour I will write the count this is number of piece, number of pieces maybe this I call it 2 4 2 4 1 3 and 3.

So, there is nothing called as piece number pieces is 1.5 can I have 1.0 number of customers can I have 1.67 number of customer whatever that is I can even have 1 customers or 2 customers ok, I can even have 1 worker or 2 workers. So, there is nothing in between that is why we do not connect this, so that is why this just bar graph is used ok.

Here the histogram thus used because histogram is actually connected to each other, if I draw them as a histogram it will be connected like this. So, it will be connected like this because, the space in between this value if between 1 and 2 there is some value let me say this value 1.5 here, it can mean something 1.5 centimetres mean something 1.5 number of piece does not mean anything.

So, this is I think I am able to explain it what is the difference between discrete and continuous data and we can we have to deal with both the kind of data. Sometimes the data is the sun then the number of defects like we will when we will discuss about the quality control, we will also discuss about number of defects and maybe proportions of defect sometimes. So, discredeotize is also there that we have to deal with ok, also one thing more like I talk about the normal distribution ok, I will talk about the normal distribution or the distributions in the next lecture.

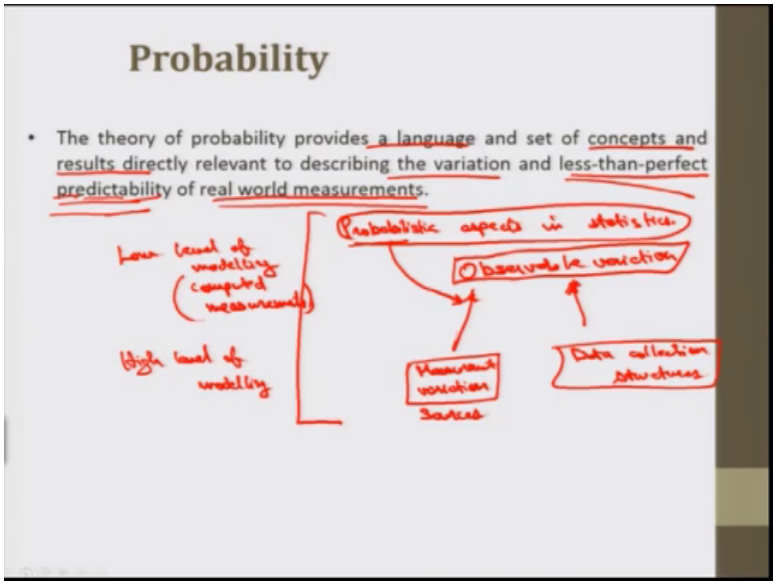
(Refer Slide Time: 14:24)




So, next is the probability, probability is a mathematical theory that is intended to describe a chance randomness or random variation ok. Probability you know probability is the chance of occurrence the chance of occurrence, if you toss a coin the probability of getting a head is 0.5 the probability of getting tail is 0.5 ok. If you toss a dies the probability of getting any number 1 2 3 4 5 or 6 is 1 by 6, if you toss 2 3 dice we can have certain permutation combination we know that this is a general probability you

know probability for A probability for B this is P A and B this is P union B the sets are there.

(Refer Slide Time: 15:20)




So, I will talk a little about probability here theory of probability provides a language and a set of concepts and results directly relevant to describing the variation and less than perfect predictability of the real world measurements. So, what will deal with in statistics is the probability, the distribution that will have would be the frequency distribution which would based upon the probability density function.

I will talk about the probability density function also, it is actually a language or the concepts or the results which are used to put our data in an organised or in a structured form. As I said the data is just unstructured when it is just data to bring make or to make this data information, it has to be put in a structured form. So, one of the way is the probability density function. So, for that probability is just a language for that. So, to it is a feasibly important it is less than perfect predictability of the real world measurements.

The real world measurements we all talk about I will talk about the confidence intervals like a just talked before, confidence interval would be 95 percent 90 percent 99 percent real world measurements are actually they are realistic and when we simulate actually we are kind of simulating while we were talking about in statistics. We what we are doing we are trying to imitate the real world problems and our result would not be perfect it will be less than perfect as said here it would be less than perfect Predictability predict.

Predict pre means before predictable predication means sometimes forecasting sometimes we just look at the sample apply our statistical tools and look at whether the sample behaves like the whole population or not for in that case probability would work. So, the subjects of probability and statistics both work together, there are certain sections statistics I will work about the probably stick aspects in statistics aspects in statistics. Now this we provide framework for the clear thinking about how sources of measurement variation and data collection structures combine to produce an observable variation.

So, we have an observer a variation which is actually drawn from the probabilistical aspects in statistics these are measurement variation is there and data collection structures are there. So, these both lead to the observable variation through the probabilistic aspects in statistics ok. So, I can even talk at in this way that the observable variation can be decomposed to quantify the important sources of measurement error measurement error various sources here.

Sources of measurement the probability is used in measurement in various forms the use of a few methods may be, a few applicability of the probability might be number one we can say the use of probability is to describe the empirical variation or uncertainty actually we use the machinery of the probability theory. We use the concept of probability theory to describe various kinds of variation and uncertainty in both measurement and collection of data in both these aspects. Then we can divide the measurement into 2 parts high level and low level of modelling low level of modelling and high level of modelling ok.

The low level is the modelling is computed the computed data is the computed compute for computed measurements actually, this is probability next I will just try to take a small break here and in the next lecture I will discuss about the frequency distributions and will talk about the probability density function, what is that and will try to plot a probability density function I will talk about the probability distributions various probability distributions which can be used to work on our data that we have collected through measurements. So, let us meet in the next lecture.

Thank you.